金融科技专利识别与分类方法构建及应用*

■ 徐璐　卢小宾　杨冠灿
中国人民大学信息资源管理学院　北京 100872

摘　要：[目的/意义]金融科技在信息和数据时代发展迅速，专利数目不断增长，同时金融科技的领域交叉和边界模糊等特性，也提升了其专利分析的难度。因此需要构建适合的识别与分类方法，从而准确高效地处理持续增长的大体量数据。[方法/过程]首先依据金融科技的内涵和功能，梳理其包含的创新类别，厘清金融科技专利的范围和边界；随后使用机器学习算法，结合文本过滤、人工判读，构建金融科技专利识别与分类的方法流程。[结果/结论]提出一种基于机器学习算法的专利识别与分类流程，能够较为准确高效地识别金融科技专利并进行分类，并通过分析所得的金融科技专利分类数据，总结当前金融科技的发展现状。

关键词：金融科技　信息经济学　专利分析　机器学习
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.11.010

伴随着大数据、云计算以及人工智能等的发展和逐渐成熟，新技术在金融行业的应用更加广泛，金融与科技的深度融合促进了金融科技(Financial Technology, FinTech)的发展，被认为是金融行业在信息时代发生的一次深刻变革与创新[1-2]。与此同时，金融科技相关的专利申请和授权数目持续增长，已经成为企业进行市场布局的重要渠道之一。因此，准确地检索金融科技专利并对其进行分析，有助于识别金融科技领域的创新，分析最具发展潜力的分支以及行业应用，从而把握金融科技的发展动态，助力国家产业和微观企业的战略性布局。

金融科技作为金融与科技的交叉领域，对其的专利识别和分析存在两大难点。一方面，金融科技所涉及的技术范围非常广泛，且目前对其的界定相对模糊，因此较难直接依据传统国际专利分类(International Patent Classification, IPC)或者确定关键词的方法进行识别和分析。另一方面，近年来随着金融科技的迅速发展，相关领域的创新成果和专利数目持续增长，因此需要构建适合的检索流程，从而能够迅速地从海量专利数据中，准确识别不同类别的金融科技创新。对此，本研究从金融科技的内涵和功能出发，通过梳理其所包含的创新类别，对金融科技专利的范围进行合理界定。同时，创新性地提出一种使用机器学习的算法，并结合文本过滤和人工判读的方法流程，进行金融科技专利识别与分类。随后，依据所得分类专利数据，分析金融科技的技术布局和发展现状。此外，本研究的方法对类似的新兴交叉领域专利分析，同样具有一定借鉴作用。

1　相关研究

1.1　金融科技相关研究

近年来学术界对金融科技的关注度持续上升，研究视角涵盖金融科技的具体技术创新与应用[3-4]、传统金融行业面临的挑战与机遇[5-6]以及相应的风险防控和市场监管[7-8]等问题，但鲜有学者从专利的角度对金融科技的发展进行探索。金融科技专利能够创造巨大的市场价值[9]，准确识别不同类型的金融科技专利，对于追踪技术发展动态、分析创新的影响因素、助力国家金融科技布局亦具有重要意义。在少数与金融科技专利相关的研究中，赵星在IPC号G06Q中，对涉及人脸识别、大数据分析预测、人工智能技术以及数字货币等方向的金融科技专利进行检

* 本文系国家自然科学基金青年项目“银行业竞争政策的实施效果及影响机制研究”(项目编号:71903189)和中国博士后科学基金项目“多元所有制银行市场中强化竞争的风险和福利效应”(项目编号:2019M660052)研究成果之一。
作者简介：徐璐(ORCID:0000-0003-2414-3180)，讲师，博士；卢小宾(ORCID:0000-0002-2866-4509)，信息分析研究中心主任，教授，博士，博士生导师，通讯作者，E-mail:luxb@ruc.edu.cn；杨冠灿(ORCID:0000-0002-1706-1884)，讲师，博士。
收稿日期:2019-12-23　修回日期:2020-03-14　本文起止页码:87-95　本文责任编辑:王传清

索和统计，指出我国金融科技领域的专利布局相对较为薄弱[10]。具有全球影响力的知识产权媒体 IPR-daily 联合 incoPat 创新指数研究中心，发布“2019 年全球金融科技发明专利排行榜（TOP100 名）”[11]，报告通过限定 IPC 号 G06Q20、G06Q30、G06Q40，针对金融、支付、购物、电子商务、保险、税务等几大应用领域，统计 2019 年全球企业公开的金融科技发明专利申请数量。上述金融科技专利的研究和实践工作，一般仅限定单个或少数几个 IPC 号进行检索，不能全面覆盖金融科技创新的所有专利，难以确保检索结果的全面性和有效性。

1.2 专利识别相关研究

传统的专利识别或者专利检索，常使用 IPC 号和关键词的检索方式。IPC 号是一种国际通用的管理和利用专利文献的工具，记载每条发明与实用新型专利信息的“分类号”和“主分类号”项，因此被广泛应用于专利的分类和检索[12-13]。一方面，使用 IPC 号，可以通过构建专利分类与产业分类的对照关系，实现快速检索和调用特定产业专利信息的目的[14]，包括专家判定、交叉检索、概率计算以及相似度等多种关联模型[15-16]。另一方面，结合 IPC 号和关键词，能够进一步优化检索方法，提高专利的识别效率[17-18]。但是，学者和实践工作者也指出，随着技术的发展，在例如智能手机和药品等众多领域，专利识别和检索都面临着跨领域、技术体系繁杂等挑战，使得基于 IPC 号的传统检索方法效果不理想[19-20]。同样，金融科技作为快速发展的新生事物，亦存在概念界定模糊和领域交叉等问题。在海量的数据中，难以直接利用 IPC 号，进行准确且规模化的专利识别与分析。因此，需要针对性地构建此类问题的专利识别的方法流程。近年来，使用机器学习算法的数据分析方法得到快速发展，被广泛应用于科学技术、经济金融等领域的问题研究[21-22]。机器学习算法能够对各类文本数据进行批量的向量化处理，与人工分析相结合，可以更高效地在海量数据中分析各类问题[23-24]，为金融科技专利的识别和分类工作提供了一种新的思路。

综上来看，已有关于金融科技的研究，多关注技术本身，探讨金融科技具体的创新场景和应用、市场影响以及政策法规等方面，但针对专利数据的研究工作较少。同时，金融科技是金融与科技交叉的新兴领域，专利数目持续增长，传统 IPC 号的检索方法难以完全涵盖金融科技的全部领域。另外，金融科技并没有明晰的概念界定，构建准确全面的关键词表亦存在困难。因此，面对海量的专利数据，本研究的主要内容有以下 3 个方面：①人工判读梳理金融科技的创新类别，厘清金融科技专利范围，作为后续金融科技专利的识别和分析的基础；②使用机器学习算法，结合文本过滤和人工判读，构建一套能够处理大体量专利数据、对金融科技专利进行识别和分类的方法流程；③结合所得的分类专利数据，分析金融科技创新的类型分布和业务应用，从而掌握金融科技发展现状。

2 金融科技专利识别和分类的流程与方法

2.1 流程框架

本研究提出一种使用机器学习、结合文本过滤和人工判读的金融科技专利识别和分类流程框架：第一，构建金融词汇列表，对涵盖金融科技必要技术的 G 和 H 类所有专利数据进行文本过滤，从而获取与金融相关的，所有可能为金融科技专利文献的标题和摘要等信息，并进行分词和 TF-IDF 特征提取。第二，依据金融科技的内涵和功能，人工梳理总结金融科技创新类别，从而界定金融科技专利的范围；进一步通过两次人工判读并结合 K-means 算法，合理构建样本集。第三，随机抽取并切分样本数据，在训练集中进行机器学习算法的参数调优；随后在测试集中进行测试，综合考虑各算法的准确率、精确率、召回率和 F1 值确定最优模型；并进行全数据集的专利识别和分类。最终构成数据采集 - 样本选择 - 模型构建的核心流程框架，如图 1 所示。图 1 中蓝色实心矩形为主要机器学习算法，橙色实心矩形为凭借金融学专业知识的人工干预和判别，绿色实心矩形为文本过滤过程。因此，在文本过滤获取专利数据的基础上，整个处理流程始终由机器学习算法所贯穿，并与人工判别进行交互，综合对金融科技专利进行识别和分类。

2.2 数据集构建及文本预处理

2.2.1 数据采集

本研究的数据来源于 lens 专利数据库，该数据库由 Cambia 和昆士兰州理工大学联合开发，提供开放性全球专利信息与学术文献数据，更新周期为 3 - 4 周，涵盖了美国专利局 2001 年以来的专利申请及 1976 年以来的专利授权的全文和图像。美国在全球金融科技发展中处于领先地位，因此本研究以美国的专利授权数据作为研究对象。本研究关注近 5 年的金融科技专利授权情况，检索时间限定在 2014 年 1 月 1 日至 2018 年 12 月 31 日，检索到全部美国授权专利共 1 528 774 条记录。

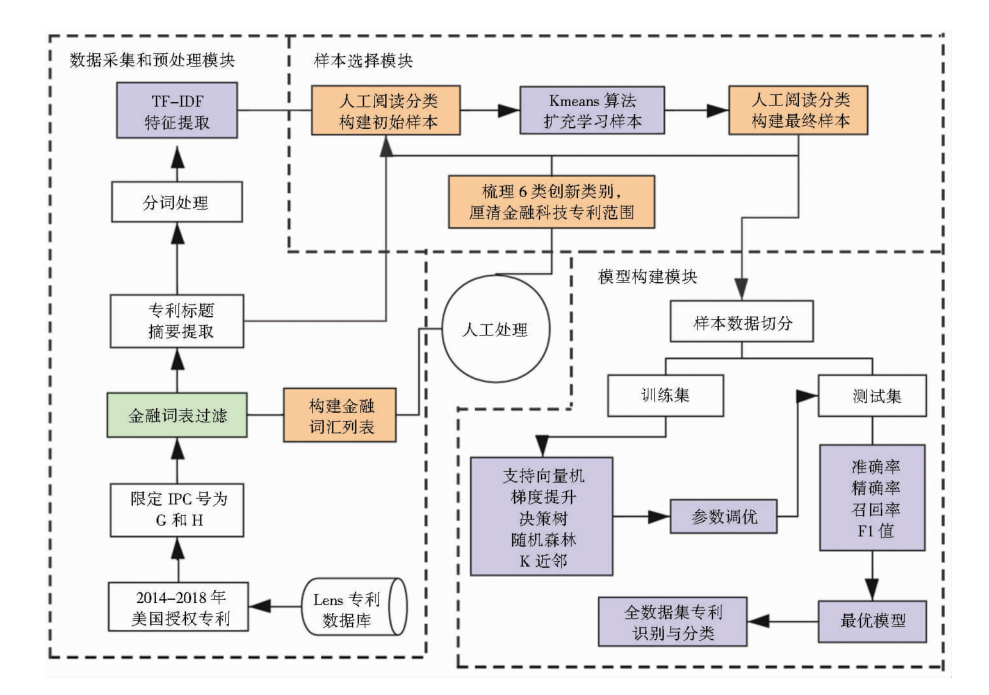


图 1　金融科技专利识别和分类方法流程

随后，为初步了解金融科技专利，本研究使用金融科技必要技术相关的术语，如"big data""cloud computing""AI"等在数据集中进行检索，结果显示国际专利分类号集中在 G(物理)和 H(电学)部类。因此，本研究进一步将 IPC 分类号限定在 G、H 两个部类，其能够涵盖金融科技所必须包含的电子计算类技术的所有专利，此过程删除了 1 328 623 条数据，剩余 200 151 条记录。

最后，本研究构建金融词汇列表，通过词表过滤的方法进一步缩小数据集。借鉴 M. Chen 等构建的金融词汇列表[9]，该列表结合 C. R. Harvey 金融词汇表及牛津金融词典中的单词及词组，提取出能够明确体现与金融服务相关联的单词(如 bourse、chargeback、futures、security、bank 等)，以及与金融服务相关的词组(如 health insurance、mutual fund 等)。在此基础上，本研究加入部分最近被确认为金融科技术语的新单词(如 digital currency、smart contract 等)，最终形成 478 个与金融服务密切相关的金融词汇表(由于篇幅所限暂且不报，备索)。使用金融词汇表，进一步过滤掉在专利文件标题和摘要中不包含表中任何金融词汇的专利，共 162 995 条数据。最终得到 37 156 条专利记录作为专利分析数据集，内容涵盖每条专利的标题、摘要和权利要求项等全部详细数据。数据集的构建过程如表 1 所示：

表 1　专利分析数据集构建

构建过程	剩余专利记录数据(条)
2014 年 1 月到 2018 年 12 月美国授权的所有专利	1 528 774
去除 IPC 号不为 G 和 H 的专利	200 151
去除标题、摘要不包含金融词汇列表中术语的专利	37 156

2.2.2　文本预处理

文本预处理环节使用 KNIME 的文本挖掘模块。首先对原始文本进行分词处理。以 OpenNLP English Tokenization 分词器为工具，以词、标点符号以及数字等进行分词，保留标点符号；之后去除标点符号、纯数字、含有异常字符、长度小于 3 的词项及常用停用词词表中的词项，如 be、the、that 等。利用 Stanford tagger 标注工具筛选名词及名词词组；最后，使用 Stanford lemmatizer 词形还原工具进一步整合相关词汇，减少英文语境下时态、语气对词项的影响，最终实现金融科技相关专利的文本处理，将专利文本语言转化为重要词项组成的词组。

随后，将文本进行向量化表示，设置特征词的数量为 3 000 维度。词袋模型是较为常用的将句子转换为向量表示的模型，该模型不考虑单词在句子中的顺序，仅考虑词表中单词在文本中出现的次数。本研究在词袋模型的基础上用词频－逆文本频率(Term Frequency - Inverse Document Frequency，TF-IDF)代替单纯的词频，进行词袋模型的构建。TF-IDF 方法是较为常用的文本特征向量化方法，用以评估词对语料库中一份文

件的重要程度。其优点是既考虑高频词的影响,同时又使用逆文本频率对词频进行加权处理,即某一单词在该文档出现的次数越多则重要性越高,但如果在越多的文档中出现,那么其在分类过程中的重要性就越小,以此反映某一单词将当前文档与其他文档区分的能力,具体计算公式如下:

$$TF(w)=\frac{\text{单词 } w \text{ 在该文档中出现的次数}}{\text{文档的单词总数}} \quad \text{公式(1)}$$

$$IDF(w)=log\left(\frac{\text{文档集中文档的总数}}{\text{包含词 } w \text{ 的文档数}+1}\right) \quad \text{公式(2)}$$

$$TF\text{-}IDF(w)=TF(w)*IDF(w) \quad \text{公式(3)}$$

2.3 金融科技专利样本集构建

2.3.1 厘清金融科技专利范围和类别

金融科技作为全新概念,其具体内涵并没有完全一致和明确的界定,如表 2 所示,不同机构和组织从各自的角度给出了金融科技的定义,但是都强调了新兴技术在金融科技领域的重要作用。与此同时,国内学者也从多维度对金融科技的内涵进行了阐述。如韩梅指出金融科技是金融和信息技术的融合型产业,内涵在于以数据和技术为驱动来提升金融行业整体运行效率,降低行业经营成本[25]。叶纯青认为金融科技是互联网金融的进一步发展,通过技术与金融的融合提升服务效率和客户体验[26]。

表 2 金融科技的内涵

年份	组织	涵义
2016	国际金融稳定理事会	前沿技术带来的金融创新,它能创造新的业务模式、应用、流程或产品,从而对金融市场、金融机构或金融服务的提供方式产生重大影响
2017	美国国家经济委员会	对付款、投资管理、募集资金、存贷款、保险、监管合规以及金融服务领域的众多活动产生影响的大范围技术创新
2017	国际证监会组织	有潜力改变金融服务行业的各种创新的商业模式以及新兴技术

对于金融科技本身及其包含的技术类型,尽管国内外组织和学者有着不同的认识,但是都普遍认同金融科技的核心是金融与科技的融合,是新技术在金融领域的应用,从而实现金融行业效率的提升和成本的节约。在此基本内涵下,前沿科技与金融不断融合的过程,催生了多种金融科技领域的技术创新。本研究所分析的金融科技专利,亦指能够应用于金融领域各项业务和流程的前沿技术性创新。

目前金融科技创新分类方法主要有两大模式:一是从金融科技在金融行业的应用方向进行划分;二是依据金融科技的底层技术进行划分。如金融稳定理事会(Financial Stability Board,FSB)在 2017 年从行业应用角度,按照支付管理、存贷款与资本筹集、投资管理、市场设施、保险五大经济功能进行分类,并进一步结合底层技术细分了物联网、电子交易、云计算、大数据、智能投顾、分布式账本、电子身份认证、移动支付等类别。

在充分了解金融科技的内涵和底层技术及其在金融领域应用的基础上,通过阅读数据集中的专利文献,本研究依据金融科技的内涵和功能,提出将金融科技专利划分为加密安全、移动支付、数据分析、物联网络、智能交易以及网络借贷这 6 个类别的技术创新,从而明确金融科技专利的内容和边界,为构建准确的专利识别和分类方法提供基础。具体分类和应用实例等如表 3 中所示。

表 3 金融科技专利分类的详细描述与具体举例

类别	描述与应用实例	具体专利
加密安全	为金融网络系统运行以及金融交易等提供安全保障的技术,应用于例如金融欺诈识别、加密货币等	US 10089683 B2:Fraud Reduction System For Transactions
移动支付	以移动终端设备为载体,借助近场通信等方式实现支付与结算的技术,应用于例如手机支付、可穿戴设备支付等	US 10083443 B1:Persistent Authentication Of A Wearable Device
数据分析	借助大数据和前沿的分析方法,解决实际金融问题的技术,应用于例如精准营销、风险管理等	US 10120923 B2:Data Discovery And Analysis Tool
物联网络	基于互联网基础的拓展网络,实现物与人的互联互通的技术,应用于例如物联网智能柜台、物联网征信等	US 10068278 B2:System And Method For Purchasing Merchandise Based On Radio Frequency Recognition
智能交易	根据客户多种交易需求,通过智能算法和产品来替代人工提供服务的技术,应用于例如智慧合约、智能投顾等	US 9721251 B1:Intelligent Capture In Mixed Fulfillment Transactions
网络借贷	借助互联网平台,以更加多样化的形式促成资金借贷双方交易的技术,应用于例如 P2P、众筹	US 8676674 B2:Peer-to-peer And Group Financial Management Systems And Methods

徐璐，卢小宾，杨冠灿.金融科技专利识别与分类方法构建及应用[J].图书情报工作,2020,64(11):87－95.

2.3.2 结合机器学习算法和人工判读构建样本集

首先,人工阅读分类,构建初始样本。使用金融科技创新企业名单[27]和金融科技高度相关的术语(如mobile transaction、mobile payment、internet security、internet of things、blockchain 等),在数据集中进行初步检索,得到765 条专利数据。表 3 概括了金融科技的 6 个具体类别,同时数据集中还包含不是金融科技的专利数据,分别记为:第 7 类,即“与金融相关但非金融科技的专利”,包括一些金融类的纯商业方法、外观设计等,这些专利虽然与金融相关,但与金融科技是“应用于金融业务和流程的前沿技术”的内涵不吻合;第 0 类,即“与金融不相关的专利”,包括一些没有具体可行的金融应用场景的、纯技术性的创新专利。因此,0－7 分类可以完全涵盖数据集中的全部类别,其中 1－6 类为金融科技专利。通过人工阅读专利文献的标题、摘要,对 765 条专利数据进行标注,初始分类结果如表 4“初始样本”列所示。

随后,使用 K-means 的方法拓展样本集。有标注样本与未标注数据集的独立同分布是进行监督学习的基本前提,即进行标注的样本应该是随机抽取且相互独立的。为保证标注样本专利的随机性,同时保证每一类有能够充分代表该类的专利数量,本研究在 765 条初始分类样本上,进一步拓展训练样本。K-means 聚类算法是一种迭代求解的聚类分析算法,通过在初始分类样本的每一类中运行 K＝1 的 K-means 算法,得到各类金融科技专利的聚类中心。进而迭代计算全体数据集中的专利与各聚类中心的余弦距离,如果一条专利的余弦距离到初始样本中某一类别的聚类中心最短,即将该专利归入该类别。利用初始分类样本,本研究使用 K-means 的方法对全部数据集进行聚类,在每类中选取离聚类中心最近的 100 篇专利,并随机抽取 100 篇专利作为补充。因此,既能够保证选取到每类的基本特征,同时增加抽取的随机性,使得对整体的代表性更好,最终得到 8 类共 1 600 条专利数据。

最后,再次人工阅读分类,确定最终样本。对1 600条样本再次进行人工阅读与分类标注,形成最终用于机器学习的专利识别与分类的样本,最终结果如表 4“最终样本”列所示。与初始分类相比,最终样本中的标注数量更丰富,对整体数据集的代表性更好。

2.4 金融科技专利识别和分类

2.4.1 训练集上进行参数调优

此阶段首先进行样本切分,随机抽取样本的 80% 作为训练集数据进行学习训练和参数调优, 剩余 20% 作为测试集数据进行验证,使用已标注的样本集寻找最优的机器学习模型。常用的机器学习分类算法包括:支持向量机(Support Vector Machine, SVM)、梯度提升决策树(Decision Tree, DT)、随机森林(Random Forest, RF)、决策树(Gradient Boosting Decision Tree, GBDT)以及 K 近邻法(K-Nearest Neighbor, KNN)等,针对不同的场景和分类问题,不同机器学习算法各自都有广泛的应用。

表 4 初始样本与最终样本中各类专利数量统计

(单位:条)

分类编号	分类名称	初始样本	最终样本
1	加密安全	114	122
2	移动支付	113	141
3	数据分析	103	57
4	物联网络	77	42
5	智能交易	87	70
6	网络借贷	48	35
7	非金融科技	109	307
0	金融不相关	114	826

选用准确率(accurate rate)、精确率(precision rate)、召回率(recall rate)和 F1 值多个指标对机器学习效果进行综合评价。准确率为所有正确分类的专利文档数目与全部专利文档数的比率,能够较为直观地衡量机器学习算法的识别和分类效果;精确率为准确分类专利文档数与所有预测为该类文档数的比率;召回率为准确分类文档数与实际文档数的比率;F1 值为同时考虑精确度和召回率的综合性评价指标。由于本研究为多分类问题,因此需要在每个二分类评价指标基础上,进一步计算整体的综合评价指标,本研究使用“宏_(macro_)”的计算方法。二分类问题和多分类问题评价指标的具体计算公式如表 5 所示。

表 5 机器学习二分类和多分类问题的评价指标

	二分类	多分类
准确率(A)	$A=\frac{TP+TN}{TP+TN+FP+FN}$	$macro_A=\frac{\sum_{i=1}^{n}TP_i}{N}$
精确率(P)	$P=\frac{TP}{TP+FP}$	$macro_P=\frac{1}{n}\sum_{i=1}^{n}P_i$
召回率(R)	$R=\frac{TP}{TP+FN}$	$macro_R=\frac{1}{n}\sum_{i=1}^{n}R_i$
F1 值	$F_1=\frac{2*P*R}{P+R}$	$macro_F_1=\frac{2*macro_P*macro_R}{macro_P+macro_R}$

因此,本研究选用所有上述 5 类机器学习算法,使用交叉验证的方法,在训练集上通过最大化综合指标 F1 值进行参数调优。最终参数选择为:支持向量机算

法取 cost = 0.6，loss = 0.1，nu = 0.5；梯度提升算法取 tree depth = 10，learning rate = 0.05；随机森林算法取 tree depth = 10；决策树算法使用 Gini index，最小描述长度(MDL)原则；K 临近取 k = 6。此时求得各机器学习方法在训练集实现的准确率、精确率、召回率和 F1 值如表 6 所示：

表 6　机器学习方法在训练集的效果

	支持向量机	梯度提升	随机森林	决策树	K 临近
准确率(A)	98.89%	99.37%	99.37%	75.73%	72.95%
精确率(P)	99.41%	99.05%	99.62%	64.17%	72.36%
召回率(R)	96.46%	98.59%	98.17%	51.37%	40.61%
F1 值	97.92%	98.82%	98.89%	57.06%	52.02%

2.4.2　测试集上确定最优模型

依据训练集所得各机器学习方法参数值设定，通过交叉验证的方法测试各类机器学习方法的效果，相同的方法可以计算出各评价指标。表 7 展示各类机器学习算法在测试集数据上的准确率、精确率、召回率和 F1 值，其中粗体显示的随机森林各项指标的综合评估效果较好，其中准确率为 75.28%，召回率为 60.27%，F1 值为 64.39%。

表 7　机器学习方法在测试集的效果

	支持向量机	梯度提升	**随机森林**	决策树	K 临近
准确率(A)	72.22%	71.94%	**75.28%**	70.83%	67.22%
精确率(P)	70.25%	67.57%	**69.10%**	57.56%	61.23%
召回率(R)	49.9%	49.16%	**60.27%**	45.00%	36.12%
F1 值	58.36%	56.92%	**64.39%**	50.51%	45.44%

随机森林是以决策树为基本分类器的一个集成学习模型，它包含多个由集成学习技术训练得到的决策树，当输入待分类的样本时，最终的分类结果由单个决策树的输出结果投票决定。随机森林克服了决策树过拟合问题，对噪声和异常值有较好的容忍性，对高维数据分类问题具有良好的可扩展性和并行性。此外，随机森林是由数据驱动的一种非参数分类方法，该算法被广泛应用到诸如生物信息、商业管理、文本分类、经济金融等实际领域。综合所得数据来看，随机森林方法对于本问题的专利识别和分类效果最好，因此本研究最终使用随机森林模型对金融科技专利进行识别与分类。

2.4.3　全数据集上专利识别与分类

利用随机森林的机器学习方法对数据集进行识别和分类。统计 2014 – 2018 年间美国的加密安全、移动支付、数据分析、物联网络、智能交易以及网络借贷 6 类金融科技专利每年的授权数目，情况如表 8 所示。结果显示，2014 年美国金融科技授权数量为 417 条，达到小高峰，2015 年有所下降，仅为 198 条，但随后呈现逐年稳定增长的趋势。2011 年金融科技的概念被正式提出[28]，金融与科技的融合逐渐由互联网金融模式走向金融科技的新模式。在此之前，依托成熟的互联网技术，各类相关专利技术爆发式产生，特别是以移动支付、智能交易和网络借贷类为代表的技术布局较快，表现在 2014 年中此三类和总体金融科技专利的授权数目较多。此后，金融科技进入纵深发展的阶段，与区块链、云计算、人工智能等底层技术的融合逐步加深，相关技术研发和布局平稳进行，因此 2015 年后各类金融科技专利逐渐呈现出平稳增长的态势，金融科技整体发展进入稳定增长的成熟期。

表 8　2014 到 2018 年间美国金融科技授权专利情况

(单位：条)

类别 / 年份	加密安全	移动支付	数据分析	物联网络	智能交易	网络借贷	总计
2014	49	147	38	27	118	38	417
2015	30	91	22	23	30	2	198
2016	34	104	30	19	19	1	207
2017	55	149	40	22	15	6	287
2018	58	191	32	24	32	11	348
总计	226	682	162	115	214	58	1 457

与此同时，所得的金融科技专利数据中共涉及 618 个 IPC 号，统计每类专利中数量排名前五的 IPC 号，如表 9 所示。结果显示，一方面，全部金融科技专利涉及的 IPC 号范围广泛；另一方面，不同类别金融科技专利的 IPC 号分布亦存在较大差异。因此，这也说明应用传统的检索方法，较难全面准确地对金融科技专利进行分析。此外，本研究的结果，也能够为未来利用 IPC 号检索和分析金融科技专利，以及构建 IPC 号与相关产业间关联等研究工作提供参考和借鉴。

表 9　各类金融科技授权专利的 IPC 分布情况

类别 / 排名	加密安全	移动支付	数据分析	物联网络	智能交易	网络借贷
1	G06Q20/40	G06Q20/32	G06Q40/00	G06K07/10	G06Q40/00	G06Q40/00
2	H04L29/06	G06Q20/40	G06F17/30	G06Q20/32	G06Q40/06	G06Q40/02
3	G06Q40/00	G06Q40/00	G06Q30/02	G06K19/07	G06Q40/04	G06Q20/10
4	G06Q20/00	G06Q20/20	H04L29/06	H04Q05/22	G06Q40/08	G06Q30/00
5	G06Q20/20	G06Q20/10	G06Q10/06	G06Q20/20	G06Q40/02	G06Q30/06

3 金融科技专利数据分析

本部分将依据所得分类专利数据，从类别分布和业务应用角度，对金融科技的发展现状进行分析和总结。

3.1 类型分布分析

按照加密安全、移动支付、数据分析、物联网络、智能交易和网络借贷6个类别对金融科技专利进行统计。如图2所示，移动支付类的金融科技专利数量最多共有682条，加密安全和智能交易紧随其后，分别有226条和214条。相对而言，网络借贷类的金融科技专利最少，仅有58条。从图3各类金融科技专利发展趋势来看，移动支付类金融科技专利占比一直保持较高水平，平均占比为47%且呈现逐年扩大的趋势，此外加密安全、数据分析和物联网络的占比也较为稳定，平均占比分别为16%、11%和8%。与此相比，智能交易类专利近年呈现明显的递减趋势，平均占比为15%，而网络借贷类专利的总体占比最小，平均仅为4%，但呈先下降后略上升的变化趋势。

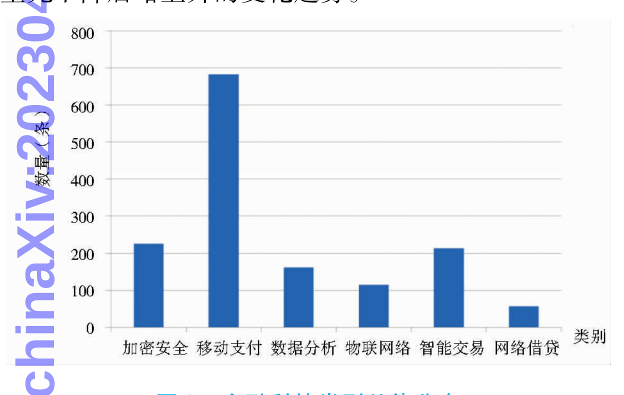


图2 金融科技类型总体分布

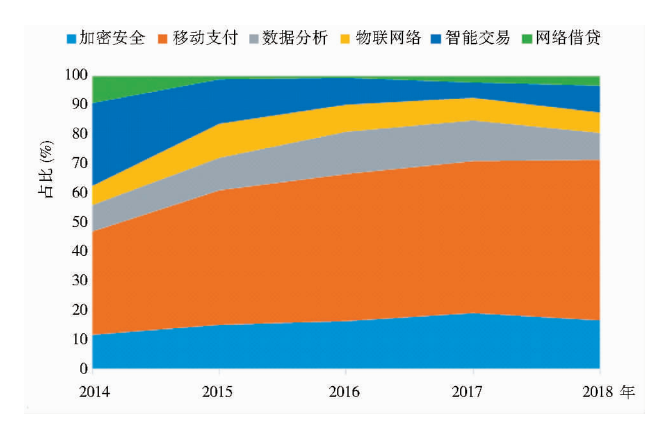


图3 各类金融科技专利变化趋势

无论从绝对数目还是发展趋势来看，移动支付类金融科技专利技术都有着绝对优势，且保持稳定增长。而智能交易类专利技术总体数量居中但逐渐缩减，网络借贷类专利技术的绝对数目和占比较少，但近年呈现增长趋势，具有一定的发展空间。未来随着5G技术、人脸识别技术等的成熟和应用，移动支付领域仍为金融科技布局的重点潜力领域，此外区块链技术的成熟也将为加密安全和网络借贷等专利布局带来新的发展机遇。

3.2 业务应用分析

金融科技专利的应用业务场景，可以划分为银行业务、资本市场业务、保险业务、支付业务、借贷业务和非特定业务，共7大类。进一步通过阅读标题、摘要并结合IPC号等信息，对所得各类金融科技的应用业务场景进行统计和分析。从不同业务应用金融科技专利总数来看（见图4），非特定业务的专利数目最多，共770条，占45.8%，其次为支付业务共460条，占27.3%，再者为银行业务，共290条，占17.2%。相比而言，金融科技专利在保险业务和借贷业务中的应用较少，分别为31条和32条，仅占总体的2%左右。

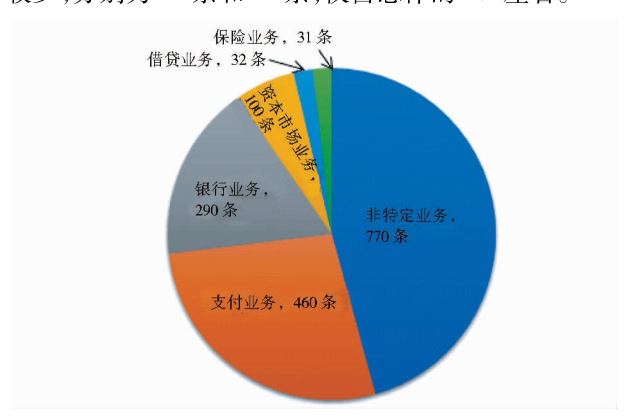


图4 金融科技专利总体的行业应用情况

从不同业务应用金融科技专利的具体类型来看，图5展示了每种业务领域内，应用数目排名前两位的专利类别。结果显示，银行业务、支付业务以及非特定业务中，应用较多的分别为移动支付和加密安全类的金融科技专利，资本市场业务和保险业务中则较多应用智能交易和数据分析类的金融科技专利，而借贷业务中移动支付和网络借贷类金融科技专利数目排名靠前。

综上，当前金融科技专利最主要的应用行业包括传统金融行业中的银行业务，以及依托互联网金融兴起的支付业务，科技与相应金融业务的结合较为深入。同时，本研究发现移动支付和加密安全类金融科技专利的应用场景最为广泛，能够在银行业务、支付业务以及其他非特定业务中提供交易支付便利和信息安全保障。相比较而言，智能交易主要应用于资本市场业务

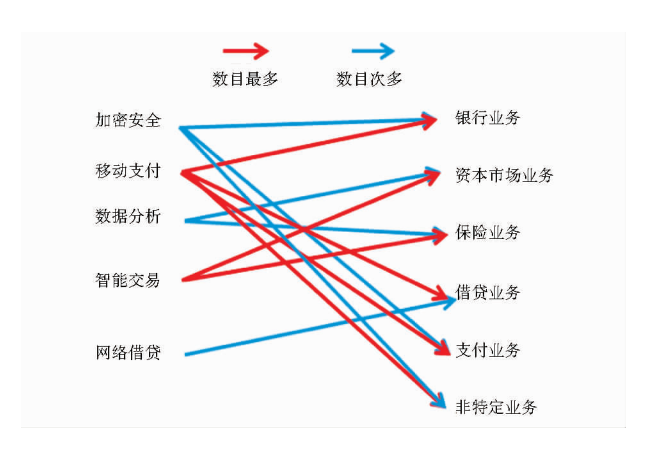


图 5　金融科技专利类别的行业应用情况

和保险业务，提供智能化的股票和保险推介；而网络借贷类专利，如“P2P”和“众筹”等相关技术，则推动了新型借贷业务的发展。

4　结语

本研究依据金融科技的内涵和功能，梳理创新的类别，从而厘清金融科技专利的范围。随后，使用机器学习算法，结合文本过滤、人工判读，构建了金融科技专利识别与分类的模型，实现自动化的专利检索。同时，针对所得分类数据，本研究亦初步探讨了金融科技专利的发展现状。面临领域交叉、技术体系繁杂和体量庞大的专利数据时，传统的专利检索方法存在检索词复杂、IPC 号检索不准确等问题。而本文提出的专利检索流程，对此类新兴交叉领域专利分析具有一定的借鉴意义。此外，同样重要的是，本研究亦能够为未来对金融科技企业创新性、产业发展影响因素等重要问题的深入探究提供可靠的专利数据及获取方法。因此，本研究具有理论和实践的双重意义。

在未来的研究工作中，还可以进一步优化文本处理的方法，提升机器学习分类模型的准确性；扩展数据集的范围，进行全球专利数据的分析，此时需要对国家间差异等方面做出更加细致的考虑。同时，后续可以进一步将金融科技的专利授权数据与公司微观数据、行业中观数据以及经济宏观数据相关联，从而深入探讨如何从公司治理和制度建设角度，促进金融科技创新的持续平稳发展。

参考文献：

[1] 牟乃密. 抢抓金融科技机遇，提升服务实体经济水平[J]. 银行家, 2018,35(6):29.

[2] 徐扬，刘姝雯，腾菲，等. 情报研究在 FinTech 实践中的应用展望[J]. 图书情报工作,2017,61(16):107－112.

[3] TAYLOR M B. The evolution of bitcoin hardware[J]. Computer, 2017, 50(9): 58－66.

[4] MOON W Y, KIM S D. Adaptive fraud detection framework for FinTech based on machine learning[J]. Advanced science letters, 2017,23(10): 10167－10171.

[5] JAPPAROVA I, RUPEIKA-APOGA R. Banking business models of the digital future: the case of Latvia[J]. European research studies, 2017,20(3): 864－878.

[6] LEE I, SHIN Y J. FinTech: ecosystem, business models, investment decisions, and challenges[J]. Business horizons, 2018,61(1): 35－46.

[7] BARBERIS J, ARNER D W. FinTech in China: from shadow banking to P2P lending [C]//TASCA P, ASTE T, PELIZZON L, et al. Banking beyond banks and money. Switzerland: Springer, 2016: 69－96.

[8] NG A W, TANG W. Regulatory risks and strategic controls in the global financial centre of China [C]//CHOI J J, POWERS M R, ZHANG X T. The political economy of Chinese finance. Bradford: Emerald Group Publishing Limited, 2016: 243－270.

[9] CHEN M A, WU Q, YANG B. How valuable is FinTech innovation? [J]. The review of financial studies, 2019, 32(5): 2062－2106.

[10] 赵星. 金融科技专利竞争态势分析 [N]. 中国知识产权报, 2018－07－25(5).

[11] IPRdaily 中文网与 incoPat 创新指数研究中心. 2019 年全球金融科技发明专利排行榜(TOP100 名)[EB/OL]. [2020－03－14]. https://www.bankbuy.net/news/detail/art_id/556.html.

[12] 宋群，吴广印. 基于 IPC 的专利数据组织初探[J]. 数字图书馆论坛,2013,9(4):67－70.

[13] 温芳芳. 专利分类号耦合分析在企业潜在合作关系识别中的应用[J]. 现代情报,2018,38(7):142－147.

[14] 彭茂祥，徐勇. 专利分类与产业分类对照关系构建及应用研究[J]. 科学管理研究,2017,35(5):30－33.

[15] 田创，赵亚娟. 一种基于相似度的专利与产业类目映射模型——以《国际专利分类》与《国民经济行业分类》为例[J]. 图书情报工作,2016,60(20):123－131.

[16] 田创，赵亚娟. 专利与产业的映射研究进展[J]. 图书情报工作, 2016,60(1):135－141.

[17] 罗晓宁，郑乃章. 科技查新中专利检索式的构建[J]. 科技情报开发与经济,2014,24(18):112－114,119.

[18] 张晨. 专利检索新策略——关键词与分类号相结合[J]. 科技情报开发与经济, 2014,24(13):112－113.

[19] 贺希佳，李文. 智能手机领域专利申请的检索技巧[J]. 中小企业管理与科技(下旬刊),2019,12(5):99－101.

[20] 谢京晶，洪丽娟. 浅谈药物领域专利申请快速检索技巧[J]. 中国发明与专利,2018,15(12):125－128.

[21] HOBERG G, PHILLIPS G. Text-based network industries and endogenous product differentiation[J]. Journal of political economy, 2016, 124(5): 1423－1465.

[22] WAGER S, ATHEY S. Estimation and inference of heterogeneous treatment effects using random forests[J]. Journal of the American Statistical Association, 2018, 113(2):1228－1242.

[23] 周源,刘宇飞,薛澜.一种基于机器学习的新兴技术识别方法:以机器人技术为例[J].情报学报,2018,37(9):939－955.
[24] VENUGOPALAN S, RAI V. Topic based classification and pattern identification in patents[J]. Technological forecasting and social change, 2015, 94(1): 236－250.
[25] 韩梅. FinTech的发展现状与金融创新分析[J]. 经济研究导刊, 2016,12(23): 88－90.
[26] 叶纯青."Fintech"与互联网金融[J]. 金融科技时代, 2016,25(8): 88.
[27] Top 100 companies in FinTech[EB/OL]. [2019－12－21]. https://www.americanbanker.com/news/top-100-companies-in-fintech-ab1071192.
[28] 谢宗晓.在金融科技时代推进金融网络安全"最佳实践"[J].中国信息安全,2017,8(7):63－66.

作者贡献说明:
徐璐:设计研究方案,金融科技分类,分析数据,撰写论文;
卢小宾:提出研究思路和框架,修改论文;
杨冠灿:构建模型,处理数据。

Identify and Classify FinTech Patent

Xu Lu　Lu Xiaobin　Yang Guancan
School of Information Resource Management, Renmin University of China, Beijing 100872

Abstract: [**Purpose/significance**] FinTech has developed rapidly in the information and data era, and the number of patents has continued to increase. At the same time, its field crossover and blurred borders characteristics have also increased the difficulty of patent analysis. Therefore, it is necessary to construct a suitable identification and classification method, so as to accurately and efficiently process the continuously growing large volume data. [**Method/process**] This paper firstly sorted out the innovation categories based on the connotation and function of FinTech, and thus clarified the scope and boundaries of FinTech patents. Then, it constructed a FinTech patent recognition and classification model based on machine learning algorithms, combining text filtering and manual interpretation. [**Result/conclusion**] This paper proposes a patent recognition and classification process, based on machine learning algorithms, which is more accurately and efficiently. By analyzing the obtained FinTech patent classification data, the research also summarizes the current FinTech development status.

Keywords: FinTech　information economics　patent analysis　machine learning